DOCUMENT RESUME

ED 219 282 SE 038 790

AUTHOR Oberlin, Lynn
TITLE What Standardized Achievement Tests Do and Do Not Tell the Teacher of Elementary and Junior High Science--The Implications for Teacher Education.
PUB DATE 2 Apr 82
NOTE 13p.; Paper presented at the Annual Meeting of the National Science Teachers Association (Chicago, IL, April 2-5, 1982).

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Elementary Education; *Elementary School Science; Junior High Schools; Science Education; *Science Tests; *Secondary School Science; *Standardized Tests; *Test Interpretation; *Test Use

ABSTRACT

Stanford Achievement Tests (Science) were examined to point out limitations of standardized achievement tests. Several factors teachers need to consider when using standardized tests were identified. Tests can be effective only if they measure the same content areas that make up the science curriculum. Good reading comprehension is needed for students to score well on science achievement tests (SATs). Finding SATs measuring the same topics as the school science program may be difficult. A well-designed, teacher-prepared test may be better than SATs which do not fit the curriculum. A series of tests (including those of former years) may be desirable. Tests from whole classes may be effective in evaluating a school science program. Scores may be improved by guessing and not leaving items blank. How a student feels on a day may affect test scores. Teaching to test or making students test-wise may raise school achievement scores but does not reflect the true condition of the science program. SATs are statistical measures based on probability and with enough data reliable inferences can be made. Overdependence on SATs can lead to designing science programs to help students score higher on the test. Tests should help teachers measure student outcome and not become curriculum determiners. (Author/JN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED219282

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

WHAT STANDARDIZED ACHIEVEMENT TESTS DO AND DO NOT TELL THE TEACHER OF ELEMENTARY AND JUNIOR HIGH SCIENCE-- THE IMPLICATIONS FOR TEACHER EDUCATION

Presented April 2, 1982

The Association for the Education of Teachers in Science

National Science Teachers Association--Chicago

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY
Lynn Oberlin
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Lynn Oberlin

Professor of Education

University of Florida

Gainesville, Florida 32611

SE038790

One purpose of this paper is to point out the limitations of standardized achievement tests. The Stanford Achievement Tests (Science) were chosen because of their wide use. Other tests could have been used with similar results. As achievement tests go, it is one of the better ones in use today.

Standardized test results are more important as a measure of student success than they have ever been before. In Elementary and Junior High School science, this means the results of a science sub-test which makes up a part of an achievement test. The series of tests used in this paper are from the Stanford Achievement Test, 1973 Edition. All science sub-tests regularly used through ninth grade level were examined. Since the Primary Level I (grade levels 1.5 to 2.4) does not have a science test, this paper will make use of the five levels which have a range of grade levels from 2.5 to 9.5. Chart 1 gives descriptive information about the science portion of these tests.

CHART 1

Stanford Achievement Tests (Science)

Test	Grade Level	Number of Items	Time in Minutes
Primary-Level II	2.5-3.4	27	20
Primary-Level III	3.5-4.4	42	25
Intermediate-Level I	4.5-5.4	60	30
Intermediate-Level II	5.5-6.9	60	30
Advanced	7.0-9.5	60	30

Primary-Level II is read to the students in a group setting. All other levels require each student to read each question and then select the correct answer. The obvious non-science skill needed here is the ability to read and understand the question. If a student can't read the question with understanding, science skills and knowledge cannot be measured. Note that in Chart 2 the intercorrelation between science and reading scores range from .55-.56 when the questions are read to the student to .80-.84 when the student must read the questions.

CHART 2

Stanford Achievement Tests (Science)

Intercorrelations-Reading, Science

Grade	Beginning of Year	End of Year
2	---	.55*
3	.56*	.83
4	.85	.81
5	.81	.80
6	.82	.81
7	.81	.83
8	.84	.83

*Primary-Level II-the questions are read to the student

Charts 3 and 4 give the number of items and percent of items, according to the publisher, for each of six different science areas.

Distribution of science areas on an achievement test is a matter of a value judgment. There is no absolute but there are some things in this distribution on the Stanford tests that should be noted. Concepts of Matter, Changes in Physical Environment, and Basic Skills of Science vary ten percent or more of the total content between various levels. This variation does not show an increase or decrease with age or grade level but just seems to zig zag back and forth. The percent of content included in Form and Functions of Living Things shows a slight increase in higher grades while Environmental Interaction questions decrease in percent of content from 22.2 to 6.7 between the Primary-Level II and the Advanced level test.

CHART 3

Content Groupings of Stanford Achievement Tests (Science)

Number of Items

Content Groupings	Number of Items				
	Primary Level II	Primary Level III	Intermediate-Level I	Intermediate-Level II	Adv.
Concepts of Energy	8	10	9	7	8
Concepts of Matter			5	13	7
Changes in Physical Environment	5	4	13	10	9
Form and Function of Living Things	6	9	14	15	17
Environmental Interaction	6	8	6	6	4
Basic Skills of Science	2	11	13	9	15
TOTAL	27	42	60	60	60

CHART 4

Content Groupings of Stanford Achievement Tests (Science)

Percent of Items

Content Groupings	Percent of Items				
	Primary Level II	Primary Level III	Intermediate-Level I	Intermediate-Level II	Adv.
Concepts of Energy	29.6	23.8	15	11.6	13.3
Concepts of Matter			8.3	21.7	11.6
Changes in Physical Environment	18.5	9.5	21.7	16.7	15
Form and Function of Living Things	22.2	21.4	23.3	25	28.3
Environmental Interaction	22.2	19	10	10	6.7
Basic Skills of Science	7.4	26.2	21.7	15	25

Looking at individual test items involves some subjective judgment. It would seem that many items could be answered correctly by an intelligent student who was a good reader and who had no knowledge or experience in the area of science. Below are listed a few such examples.

Primary Level III

"#38. One thermometer is put into a gallon of water. Another thermometer is put into a pint of water. Both thermometers read 60°F.

Which statement is correct?

5. The gallon of water is warmer than the pint of water.
6. Both the gallon and the pint of water are the same temperature.
7. The pint of water is warmer than the gallon of water.
8. The thermometers are probably wrong."

Intermediate Level I

"#13.

If you arrange these thermometers from the lowest temperature to the highest temperature, the order would be--

1 B, C. A 3 C, A, B
2 A, B, C 4 A, C, B"

Advanced

"The chart shows the data recorded from observations of a stone falling under the action of its weight alone.

Time (sec.)	Distance (ft.)	Speed (ft./sec.)
0	0	0
1	16	32
2	64	64
3	144	96
4	256	128
5	400	?

6. At the end of 1 second, how many feet had the stone fallen?
   5 64 6 0 7 16 8 32

7. When the stone had fallen a distance of 144 feet, at how many feet per second was it falling?
   1 144 2 96 3 128 4 160

8. It takes 5 seconds for the stone to fall a distance of 400 feet. How many feet per second will the stone probably be traveling then?
   5 160 6 128 7 400 8 144"

Most questions were well constructed as you would expect from the group of professionals who wrote this test. There are a few items, however, which deserve some attention. These are listed below.

Primary II Level II

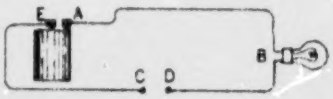
Read to Students

"#23. Look at the thermometer in the box; then look at the next three pictures. Which scene would you expect to see outside your window?"

The answer keyed as correct is the middle picture. If you lived in Northern Florida this last winter, the correct answer would be the one on the left.

Intermediate Level II

"Questions 25 and 26 are based on the picture below.



25. The set-up in the above diagram can be used to find out whether an electric current will flow through an object. To do this, you would place the object so that it would contact--
    1 B and D 3 A and B
    2 C and D 4 E and C

26. In the experiment diagrammed above, a spoon was placed across the contacts C and D. The light bulb lit. Which of the following would be the best conclusion?

    5. The spoon is made of metal.
    6. Batteries produce electricity.
    7. Electric current flowed through the spoon.
    8. The light bulb is not good."

As you can see, the correct answer to question 25 is given to the student in question 26.

"#45. Green plants are put into an aquarium because the green plants--

1. Make the aquarium look pretty.
2. Make the oxygen which is used by the animals.
3. Have leaves.
4. Grow into larger plants."

The answer keyed as correct is "2" although in many cases the answer which is really correct may be "1".

On a multiple-choice test it is always possible to mark an answer to every question even if you have no idea what the right answer should be. Given just the sheet to mark answers on, a student could randomly mark one answer to each question. If each question had four choices, the student has one chance in four of choosing the correct answer. Given to a large enough group, the average score on such a test would be one-fourth of the number of items. Chart 5 shows average random change scores and the corresponding grade equivalents of such a score.

CHART 5

Test	Grade Levels	Number of Items	Average Random Chance Score	Grade Equivalents
Primary I	2.5-3.4	27	9	1.1
Primary II	3.5-4.4	42	11	2.0
Intermediate I	4.5-5.4	60	15	2.9
Intermediate II	5.5-6.9	60	15	3.4
Advanced	7.0-9.5	60	15	4.2

It should be noted that these are averages and many of the students would be above and below these averages.

The effect of guessing is not really this simple to explain. Most students do know the answers to some questions and guess on the rest. An

example of this would be a seventh grader who knew the answer to 20 questions and guessed on the other 40. Figuring average luck or random chance, his score would be 30. In terms of grade equivalents, his guessing raised him from 5.7 to 8.1.

In actual testing conditions, students usually do know some choices are wrong which then increases their chances. In the example above let's say that out of the 40 questions for which the student did not know the answer, there were 20 of them in which he knew two of the choices could not be right. That gives him 20 answers he knew, 20 that his chances were one out of two, and 20 that his chances were one out of four. The average of scores from these conditions would be 35. His grade equivalent now would have gone from 5.7 to 9.4 by intelligent guessing. To increase the score of a student be sure he leaves no answer space blank and eliminates answers he knows are wrong. This is called being test-wise!

An individual student's score may be interpreted by using the standard error of measurement. A student's test score is an approximation of his true score. Two times out of three the true score will be within (+ or -) one standard error of the test score. For more precision, 19 times out of 20 the true score will be within (+ or -) two standard errors of the test score. Chart 6 shows standard errors of measurement in terms of grade equivalents.

CHART 6

Stanford Achievement Tests (Science)

Standard Error of Measurement in Grade Equivalents

	Beginning of Year		End of Year	
Grade	Test	S.E. G.E.	Test	S.E. G.E.
2			Primary II	.7
3	Primary II	.7	Primary III	.4
4	Primary III	.4	Intermediate I	.5
5	Intermediate I	.4	Intermediate II	.6
6	Intermediate II	.6	Intermediate II	.6
7	Advanced	.9	Advanced	.8
8	Advanced	.8	Advanced	.8

An example of this would be an end of the year fifth grade student scoring a 5.0 grade level. There would be two chances out of three that his true score would be from 4.4 to 5.6 and 19 chances in 20 that it would be from 3.8 to 6.2.

Factors Science Teachers Need to Consider:

1. Tests can be effective only if they measure the same broad areas of content that make up the science curriculum. To teach one thing and measure a different one does not make sense.
2. Good reading comprehension is needed for a student to score well on a science achievement test. If he can't read it, he can't answer it correctly.
3. Finding a science achievement test which measures the same topics as the school science program may be difficult as there is no accepted order of topics as in reading or mathematics. Different programs present the same content at different levels.

4. A well designed teacher-made test may be better than a standardized test which does not fit the curriculum.
5. While one test of a single student leaves a lot of room for doubt, a series of tests may be helpful in getting at the truth. This can include tests of former years.
6. Tests from the whole class can be effective in evaluating the school science program.
7. Students may improve their scores by guessing and not leaving an answer blank on an achievement test. A blank answer is always wrong. Tests at elementary and junior high school level are not corrected for guessing but are scored only on the number correct.
8. How a student feels on testing day may affect his score. Sick students generally score lower.
9. If the school wishes to raise achievement test scores, it may do so by teaching to test areas and training the students to be test-wise. Such test gains do not reflect the true condition of the science program.
10. Standardized tests are valid only when used under standardized conditions. The directions should be read to students and time limits followed precisely.
11. Standardized achievement tests are statistical measures based on probability. Certainty is impossible but with enough data reliable inferences can be made.
12. Over dependence on standardized tests can lead to designing the science program to help students score higher on the test. Tests should help teachers to measure student outcome and not become curriculum determiners.

Implications for Science Teacher Education

Achievement tests are a fact of life in today's schools. It is very important that science teachers are knowledgeable about tests and testing. They must know just what a test can and can not tell them about a student. This includes knowing the limitations of a test, how to choose a test, and how to construct a test for areas not covered by standardized tests.

Science teachers must be curriculum planners and take an active part in the development of science programs. They must resist the trend to plan programs to help students score higher on tests and insist that the test must accommodate the program, not the program accommodate the test. Science teachers must recognize that a good science program may include some things that are not on today's science achievement tests.